A BYY Split-and-Merge EM Algorithm for Gaussian Mixture Learning

Lei Li and Jinwen Ma*

Department of Information Science, School of Mathematical
Sciences and LAMA, Peking University, Beijing, 100871, China
jwma@math.pku.edu.cn

Abstract. Gaussian mixture is a powerful statistic tool and has been widely used in the fields of information processing and data analysis. However, its model selection, i.e., the selection of number of Gaussians in the mixture, is still a difficult problem. Fortunately, the new established Bayesian Ying-Yang (BYY) harmony function becomes an efficient criterion for model selection on the Gaussian mixture modeling. In this paper, we propose a BYY split-and-merge EM algorithm for Gaussian mixture to maximize the BYY harmony function by splitting or merging the unsuited Gaussians in the estimated mixture obtained from the EM algorithm in each time dynamically. It is demonstrated well by the experiments that this BYY split-and-merge EM algorithm can make both model selection and parameter estimation efficiently for the Gaussian mixture modeling.

Keywords: Bayesian Ying-Yang (BYY) harmony learning, Gaussian mixture, EM algorithm, Model selection, Parameter estimation.

1 Introduction

As a powerful statistical tool, Gaussian mixture has been widely used in the fields of information processing and data analysis. Generally, the parameters of Gaussian mixture can be estimated by the expectation-maximization (EM) algorithm [1] under the maximum-likelihood framework. However, the EM algorithm not only suffers from the problem of local optimum, but also converges to a wrong result in the situation that the actual number of Gaussians in the mixture is set incorrectly. Since the number of Gaussians is just the scale of the Gaussian mixture model, the selection of number of Gausians in the mixture is also referred to as the model selection.

In a conventional way, we can choose a best number k^* of Gaussians via some selection criterion, such as Akaike's information criterion (AIC) [2] and the Bayesian inference criterion [3]. However, these criteria have certain limitations and often lead to a wrong result. Moreover, this approach involves a large computational cost since the entire process of parameter estimation has to be repeated for a number of different choices of k.

In past several years, with the development of the Bayesian Ying-Yang (BYY) harmony learning system and theory [4,5], a new kind of BYY harmony learning

* Corresponding author.

F. Sun et al. (Eds.): ISNN 2008, Part I, LNCS 5263, pp. 600–609, 2008.

algorithms, such as the adaptive, conjugate, natural gradient, simulated annealing and fixed-point learning algorithms [6,7,8,9,10], have been established to make model selection automatically during the parameter learning. Although these new algorithms are quite efficient for both model selection and parameter estimation for the Gaussian mixture modeling, they must satisfy a particular assumption that k is larger than the number of actual Gaussians in the sample data, but not too much. Actually, if k is too larger than the true one, these algorithms often converge to a wrong result. Nevertheless, how to overestimate the true number of Gaussians in the sample data in such a way is also a difficult problem.

In this paper, we propose a new kind of split-and-merge EM algorithm that maximizes the harmony function gradually in each time through the split-and merge operation on the estimated mixture from the EM algorithm and terminates at the maximum of the harmony function. Since the maximization of the harmony function corresponds to the correct model selection on the Gaussian mixture modeling [11] and the split-and-merge operation can escape from a local maximum of the likelihood function, the BYY split-and-merge EM algorithm can lead to a better solution for both model selection and parameter estimation.

The rest of the paper is organized as follows. In Section 2, we revisit the EM algorithm for Gaussian mixtures. We further introduce the BYY learning system and the harmony function in Section 3. In Section 4, we present the BYY split-and-merge EM algorithm. Several experiments on the synthetic and real-world data sets, including a practical application of unsupervised color image segmentation, are conducted in Section 5 to demonstrate the efficiency of the proposed algorithm. Finally, we conclude briefly in Section 6.

2 The EM Algorithm for Gaussian Mixtures

The probability density of the Gaussian mixture of k components in $\Re^d$ can be described as follows:

$$\Phi(x) = \sum_{i=1}^{k} \pi_i \phi(x|\theta_i), \qquad \forall x \in \Re^d, \tag{1}$$

where $\phi(x|\theta_i)$ is a Gaussian probability density with the parameters $\theta_i = (m_i, \Sigma_i)$ (m_i is the mean vector and Σ_j is the covariance matrix which is assumed positive definite) given by

$$\phi(x|\theta_i) = \phi(x|m_i, \Sigma_i) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma_i|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-m_i)^\top \Sigma_i^{-1}(x-m_i)}, \tag{2}$$

and $\pi_i \in [0,1] (i = 1, 2, \cdots, k)$ are the mixing proportions under the constraint $\sum_{i=1}^{k} \pi_i = 1$. If we encapsulate all the parameters into one vector: $\Theta_k = (\pi_1, \pi_2, \ldots, \pi_k, \theta_1, \theta_2, \ldots, \theta_k)$, then, according to Eq.(1), the density of Gaussian mixture can be rewritten as:

$$\Phi(x|\Theta_k) = \sum_{i=1}^{k} \pi_i \phi(x|\theta_i) = \sum_{i=1}^{k} \pi_i \phi(x|m_i, \Sigma_i). \tag{3}$$

For the Gaussian mixture modeling, there are many learning algorithms. But the EM algorithm may be the most well-known one. By alternatively implementing the E-step to estimate the probability distribution of the unobservable random variable and the M-step to increase the log-likelihood function, the EM algorithm can finally lead to a local maximum of the log-likelihood function of the model. For the Gaussian mixture model, given a sample data set $\mathcal{S} = \{x_1, x_2, \cdots, x_N\}$ as a special incomplete data set, the log-likelihood function can be expressed as follows:

$$\log p(\mathcal{S} \mid \Theta_k) = \log \prod_{t=1}^{N} \phi(x_t \mid \Theta_k) = \sum_{t=1}^{N} \log \sum_{i=1}^{k} \pi_i \phi(x_t \mid \theta_i), \tag{4}$$

which can be optimized iteratively via the EM algorithm as follows:

$$P(j|x_t) = \frac{\pi_j \phi(x_t \mid \theta_j)}{\sum_{i=1}^{k} \pi_i \phi(x_t \mid \theta_i)}, \tag{5}$$

$$\pi_j^+ = \frac{1}{N} \sum_{t=1}^{N} P(j|x_t), \tag{6}$$

$$\mu_j^+ = \frac{1}{\sum_{t=1}^{N} P(j|x_t)} \sum_{t=1}^{N} P(j|x_t) x_t, \tag{7}$$

$$\Sigma_j^+ = \frac{1}{\sum_{t=1}^{N} P(j|x_t)} \sum_{t=1}^{N} P(j|x_t)(x_t - \mu_j^+)(x_t - \mu_j^+)^T. \tag{8}$$

Although the EM algorithm can have some good convergence properties in certain situations ([12,13,14]), it certainly has no ability to determine the proper number of the components for a sample data set because it is based on the maximization of the likelihood. In order to overcome this weakness, we will utilize the BYY harmony function as the criterion for the Gaussian mixture modeling.

3 BYY Learning System and Harmony Function

In a BYY learning system, each observation $x \in X \subset \mathcal{R}^d$ and its corresponding inner representation $y \in Y \subset \mathcal{R}^m$ are described with two types of Bayesian decomposition $p(x, y) = p(x)p(y|x)$ and $q(x, y) = q(y)q(x|y)$, which are called them Yang and Ying machine respectively. For the Gaussian mixture modeling, y is limited to be an integer in $Y = \{1, 2, \ldots, k\}$. With a sample data set $\mathcal{D}_x = \{x_t\}_{t=1}^{N}$, the aim of the BYY learning system is to specify all the aspects of $p(y|x)$,$p(x)$,$q(x|y)$,$q(y)$ by maximizing the following harmony functional:

$$H(p \parallel q) = \int p(y \mid x)p(x) \ln[q(x \mid y)q(y)]dxdy - \ln z_q, \tag{9}$$

where z_q is a regularization term and will often be neglected.

If both $p(y \mid x)$ and $q(x \mid y)$ are parametric, i.e, from a family of probability densities with a parameter $\theta \in R^d$, the BYY learning system is called to have a

Bi-directional Architecture (BI-Architecture). For the Gaussian mixture modeling, we use the following specific BI-Architecture of the BYY learning system. $q(j) = \alpha_j(\alpha_j \geq 0$ and $\sum_{j=1}^{k} \alpha_j = 1)$ and $p(x) = \frac{1}{N}\sum_{t=1}^{N} \delta(x-x_t)$. Furthermore, the BI-architecture is constructed with the following parametric forms:

$$p(y = j \mid x) = \frac{\alpha_j q(x \mid \theta_j)}{q(x \mid \Theta_k)}, \qquad q(x \mid \Theta_k) = \sum_{j=1}^{k} \alpha_j q(x \mid \theta_j) \tag{10}$$

where $q(x \mid \theta_j) = q(x \mid y = j)$ with θ_j consisting of all its parameters and $\Theta_k = \{\alpha_j, \theta_j\}_{j=1}^{k}$. Substituting all these component densities into Eq.(9), we have the following harmony function:

$$H(p \parallel q) = J(\Theta_k) = \frac{1}{N}\sum_{t=1}^{N}\sum_{j=1}^{k} \frac{\alpha_j q(x_t \mid \theta_j)}{\sum_{i=1}^{k} \alpha_i q(x_t \mid \theta_j)} ln[\alpha_j q(x_t \mid \theta_j)]. \tag{11}$$

When each $q(x \mid \theta_j)$ is a Gaussian probability density given by Eq.(2), $J(\Theta_k)$ becomes a harmony function on Gaussian mixtures. Furthermore, it has been demonstrated by the experiments [6,7,8,9,10] and theoretical analysis [11] that as this harmony function arrives at the global maximization, a number of Gaussians will match the actual Gaussians in the sample data, respectively, with the mixing proportions of the extra Gaussians attenuating to zero. Thus, we can use the harmony function as the reasonable criterion for model selection on Gaussian mixture.

4 The BYY Split-and-Merge EM Algorithm

With the above preparations, we begin to present our BYY split-and-merge EM algorithm. Given a sample data set $\mathcal{S}$ from an original mixture with $k^*(> 1)$ actual Gaussians, we use the EM algorithm to get k estimated Gaussians with the initial parameters. If $k \neq k^*$, some estimated Gaussians cannot match the actual Gaussans properly and it is usually efficient to utilize a split-and-merge EM algorithm to split or merge those unsuited Gaussians dynamically. Actually, the main mechanisms of the split-and-merge EM algorithm are the split and merge criteria. Based on the BYY harmony function and the analysis of the overlap between two Gaussians in a sample data set, we can construct the split and merge criteria as well as the split-and-merge EM algorithm in the following three subsections.

4.1 The Harmony Split Criterion

After each usual EM procedure, we get the estimated parameters Θ_k in the Gaussian mixture. According to Eq.(11), the harmony function $J(\Theta_k)$ can be further expressed in the sum form: $J(\Theta_k) = \sum_{j=1}^{k} H_j(p_j \parallel q_j)$, where

$$H(p_j \parallel q_j) = \frac{1}{N}\sum_{t=1}^{N} \frac{\alpha_j q(x_t \mid \theta_j)}{\sum_{i=1}^{k} \alpha_i q(x_t \mid \theta_j)} ln[\alpha_j q(x_t \mid \theta_j)]. \tag{12}$$

Clearly, $H(p_j \parallel q_j)$ denotes the harmony or matching level of the $j-th$ estimated Gaussian with respect to the corresponding actual Gaussian in the sample data set. In order to improve the total harmony function, we can split the Gaussian with the least component harmony value $H(p_j \parallel q_j)$. That is, if $H(p_r \parallel q_r)$ is the least one, the harmony split criterion will implement the split operation on the $r-th$ estimated Gaussian. Specifically, we divide it into two components i', j' with their parameters being designed as follows (refer to [15]).

Generally, the covariance matrix Σ_r can be decomposed as $\Sigma_r = USV^T$, where $S = diag[s_1, s_2, \cdots, s_d]$ is a diagonal matrix with nonnegative diagonal elements in a descent order, U and V are two (standard) orthogonal matrices. Then, we further set $A = U\sqrt{S} = Udiag[\sqrt{s_1}, \sqrt{s_2}, \cdots, \sqrt{s_d}]$ and get the first column A_1 of A. Finally, we have the parameters for the two split Gaussians as follows, where γ, μ, β are all set to be 0.5.

$$\alpha_{i'} = \gamma\alpha_r, \alpha_{j'} = (1-\gamma)\alpha_r; \tag{13}$$

$$m_{i'} = m_r - (\alpha_{j'}/\alpha_{i'})^{1/2}\mu A_1; \tag{14}$$

$$m_{j'} = m_r + (\alpha_{i'}/\alpha_{j'})^{1/2}\mu A_1; \tag{15}$$

$$\Sigma_{i'} = (\alpha_{j'}/\alpha_{i'})\Sigma_r + ((\beta - \beta\mu^2 - 1)(\alpha_r/\alpha_{i'}) + 1)A_1A_1^T; \tag{16}$$

$$\Sigma_{j'} = (\alpha_{i'}/\alpha_{j'})\Sigma_r + ((\beta\mu^2 - \beta - \mu^2)(\alpha_r/\alpha_{j'}) + 1)A_1A_1^T. \tag{17}$$

4.2 The Overlap Merge Criterion

For the $r-th$ component with the sample x_t, we introduce a special function: $U(x_t, r) = p(y = r \mid x_t)(1 - p(y = r \mid x_t))$, where $p(y = r \mid x_t)$ is just the posterior probability of the sample x_t over the $r-th$ component. Clearly, in the estimated Gassians mixture, $U(x_t, r)$ is a special measure of the degree of the sample x_t belonging to the $r-th$ component. With this special measure, we can define the degree of the overlap between two components under a given sample data set $\mathcal{S}$ as follows:

$$F_{i,j} = \frac{\sum_{\Omega_j^\varepsilon} U(x_t, i) * \sum_{\Omega_i^\varepsilon} U(x_t, j)}{\#\Omega_i^\varepsilon * \#\Omega_j^\varepsilon * dist(i,j)} \tag{18}$$

where $\Omega_r^\varepsilon = \{x_t | p(y = r \mid x_t) > 0.5 \& U(x_t, r) \geq \varepsilon\}$ and $dist(i,j)$ is the Mahalanobis distance between $i-th$ and $j-th$ components.

Since $F_{i,j}$ is a measure of overlap between components i and j, it is clear that the two components should be merged together if $F_{i,j}$ is large enough. Thus, the overlap merge criterion is that if $F_{i,j}$ is the highest one, the $i-th$ and $j-th$ components will be merged into one component by the following rules ([15]):

$$\alpha_r = \alpha_i + \alpha_j; \tag{19}$$

$$m_r = \alpha_i m_i + \alpha_j m_j; \tag{20}$$

$$\Sigma_r = (\alpha_i\Sigma_i + \alpha_j\Sigma_j + \alpha_i m_i m_i^\intercal + \alpha_j m_j m_j^\intercal - \alpha_r m_r m_r^\intercal)/\alpha_r. \tag{21}$$

4.3 Procedure of the BYY Split-and-Merge EM Algorithm

With the harmony split criterion and the overlap merge criterion, we can present the procedure of the BYY split-and-merge EM algorithm as follows:

1. According to the initial values of k and the parameters Θ_k, implement the usual EM algorithm and then compute $J(\Theta_k)$.
2. Implement the following split and merge operations independently.

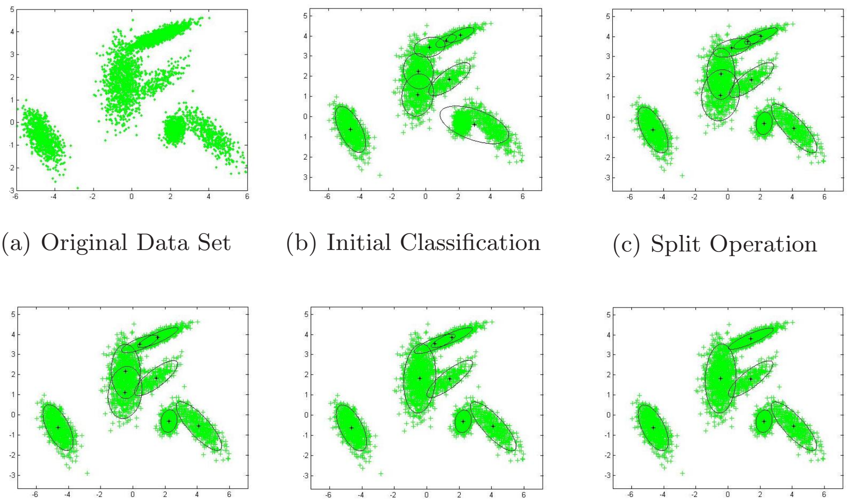
   Split Operation: With the current k and the obtained parameters Θ_k, split the Gaussian $q(x|\theta_r)$ of the least component harmony value into two new Gaussians $q(x|\theta'_j)$ and $q(x|\theta''_j)$ according to Eqs.(13)-(17). Then, implement the usual EM algorithm from the parameters of the previous and split Gaussians to obtain the updated parameters Θ_{split} for the current mixture of $k+1$ Gaussians; compute $J(\Theta_{split})$ on the sample data set and denote it by J_{split}.

   Merge Operation: With the current k and the parameters Θ_k, merge the two Gaussians with the highest degree of overlap into one Gaussian according to Eqs.(19)-(21) and implement the usual EM algorithm from the parameters of the previous and merge Gaussians to obtain the updated parameters Θ_{merge} for the current mixture of $k-1$ Gaussians; compute $J(\Theta_{merge})$ on the sample data set and denote it by J_{merge}.
3. Compare the three value $J_{old} = J(\Theta_k)$, J_{split} and J_{merge} and continue the iteration until stop.
   (i). If $J_{split} = max(J_{old}, J_{split}, J_{merge})$, we accept the result of the split operation and set $k = k+1, \Theta_{k+1} = \Theta_{split}$, go to step 2;
   (ii). If $J_{merge} = max(J_{old}, J_{split}, J_{merge})$, we accept the result of the merge operation and set $k = k-1, \Theta_{k-1} = \Theta_{merge}$, go to step 2;
   (iii). If $J_{old} = max(J_{old}, J_{split}, J_{merge})$, we stop the algorithm with the current Θ_k as the final result of the algorithm.

It can be easily found from the above procedure that both the split and merge operations try to increase the total harmony function and the stopping criterion tries to prevent from splitting and merging too many Gaussians. Thus, the harmony function criterion will make a correct model selection, while the usual EM algorithm still maintains a maximum likelihood (ML) solution of the parameters Θ_k. Therefore, this split-and-merge EM procedure will lead to a better solution on the Gaussian mixture modeling for both model selection and parameter estimation.

5 Experimental Results

In this section, we demonstrate the BYY split-and-merge EM algorithm through a simulation experiment and two applications for the classification of two real-world datasets and unsupervised color image segmentation. Moreover, we compare it with the greedy EM algorithm given in [16] on unsupervised color image segmentation.



(a) Original Data Set (b) Initial Classification (c) Split Operation

(d) Merge Operation (e) Merge Operation (f) Final Result

Fig. 1. (a): The Synthetic Data Set with Six Gaussians Used in the Simulation Experiment. (b)-(e): The Experimental Results at the Four Typical Iterations of the BYY Split-and-Merge EM Algorithm. (f). The Final Experimental Result of the BYY Split-and-Merge EM Algorithm.

5.1 Simulation Result

In the simulation experiment, a synthetic data set containing six bivariate Gaussian distributions (i.e. $d = 2$) with certain degree of overlap, which is shown in Fig. 1(a), was used to demonstrate the performance of the BYY split-and-merge EM algorithm. The initial mean vectors were obtained by the k-means algorithm at $k = 8$, which is shown in Fig.1(b). The BYY split-and-merge EM algorithm was implemented on the synthetic data set until $J(\Theta_k)$ arrived at a maximum. The typical results during the procedure of the BYY split-and-merge EM algorithm are shown in Fig.1(c)-(f), respectively. It can be observed from these figures that the BYY split-and-merge EM algorithm not only detected a correct number of Gaussians for the synthetic data set, but but also led to a good estimation of the parameters in the original Gaussian mixture.

5.2 On Classification of the Real-World Data

We further applied the BYY split-and-merge EM algorithm to the classification of the Iris data (3-class, 4-dimensional, 150 samples) and the Wine data (3-class, 13-dimensional, 178 samples). In the both experiments, we masked the class indexes of these samples and used them to check the classification accuracy of the BYY split-and-merge EM algorithm. For quick convergence of the algorithm, a low threshold T is set such that as long as some mixing proportion was less than

Table 1. The Classification Results of the BYY Split-and-Merge EM Algorithm on Real-world Data Sets

The data set	ε	T	k	The classification accuracy
Iris data set	0.2	0.10	2	98.0% ±0.006
Wine data set	0.2	0.10	4	96.4% ±0.022

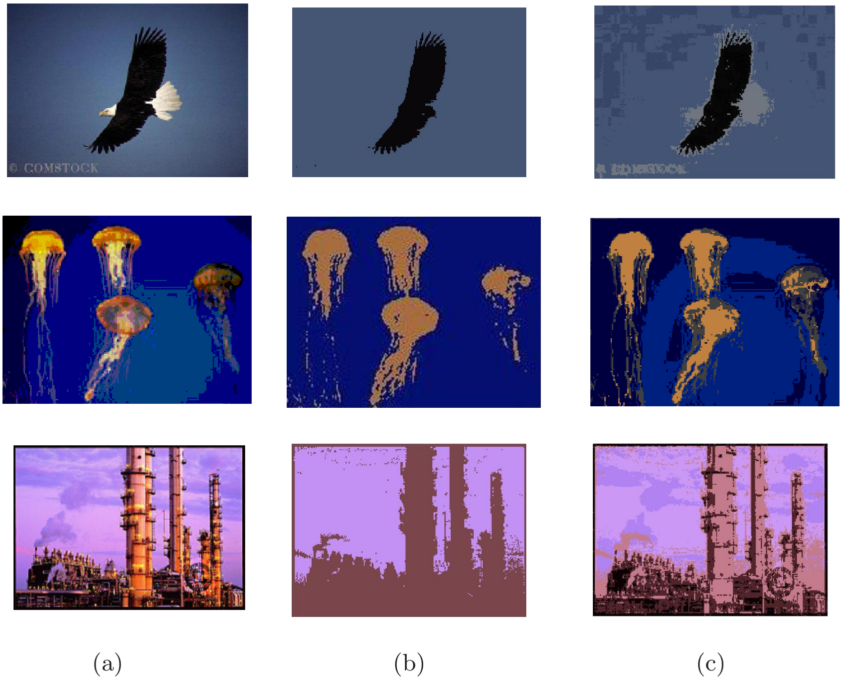


Fig. 2. The Experimental Results on Unsupervised Color Image Segmentation. (a). The Original Color Images. (b). The Segmentation Results of the BYY Split-and-Merge EM Algorithm. (c). The Segmentation Results of the Greedy EM Algorithm.

T, the corresponding Gaussian in the mixture would be discarded immediately. In the experiments, for each data set with $k = 2, 4$, we implemented the algorithm from the different initial parameters for 100 times. The classification results of the algorithm on the Iris and wine data sets are summarized in Table 1. It can be seen from Table 1 that their classification accuracies were rather high and stable (with a very small deviation from the average classification accuracy).

5.3 On Unsupervised Color Image Segmentation

Segmenting a digital color image into homogenous regions corresponding to the objects (including the background) is a fundamental problem in image processing. When the number of objects in an image is not known in advance, the image segmentation problem is in an unsupervised mode and becomes rather difficult in practice. If we consider each object as a Gaussian distribution, the whole color image can be regarded as a Gaussian mixture in the data or color space. Then, the BYY split-and-merge EM algorithm provides a new tool for solving this unsupervised color image segmentation problem. Actually, we applied it to the unsupervised color image segmentation on three typical color images that are expressed in the three-dimensional color space by the RGB system and also compared it with the greedy EM algorithm.

The three color images for the experiments are given in Fig. 2(a). The segmentation results of these color images by the BYY split-and-merge EM algorithm are given in Fig.2(b). For comparison, the segmentation results of these color images by the Greedy EM algorithm are also given in Fig. 2(c). From the segmented images of the two algorithms given in Fig. 2, it can be found that the BYY split-and-merge EM algorithm could divide the objects from the background efficiently. Moreover, our proposed algorithm could obtain a more accurate segmentation on the contours of the objects in each image.

6 Conclusions

Under the framework of the Bayesian Ying-Yang (BYY) harmony learning system and theory, we have established a BYY split-and-merge EM algorithm with the help of the conventional EM algorithm. By splitting or merging the unsuited estimated Gaussians obtained from the EM algorithm, the BYY split-and-merge EM algorithm can increase the total harmony function at each time until the estimated Gaussians in the mixture match the actual Gaussians in the sample data set, respectively. It is demonstrated well by the simulation and practical experiments that the BYY split-and-merge EM algorithm can achieve a better solution for the Gaussian mixture modeling on both model selection and parameter estimation.

Acknowledgments. This work was supported by the Natural Science Foundation of China for grant 60771061.

References

1. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximun Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Soceity B 39, 1–38 (1977)
2. Akaike, H.: A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 19, 716–723 (1974)
3. Scharz, G.: Estimating the Dimension of a Model. The Annals of Statistics 6, 461–464 (1978)
4. Xu, L.: Best Harmony, Unified RPCL and Automated Model Selection for Unsupervised and Supervised Learning on Gaussian Mixtures, Three-layer Nets and ME-RBF-SVM Models. International Journal of Neural Systems 11, 43–69 (2001)

5. Xu, L.: BYY Harmony Learning, Structural RPCL, and Topological Self-Organzing on Mixture Modes. Neural Networks 15, 1231–1237 (2002)
6. Ma, J., Wang, T., Xu, L.: A Gradient BYY harmony Learning Rule on Gaussian Mixture with Automated Model Selection. Neurocomputing 56, 481–487 (2004)
7. Ma, J., Gao, B., Wang, Y., et al.: Conjugate and Natural Gradient Rules for BYY Harmony Learning on Gaussian Mixture with Automated Model Selection. International Journal of Pattern Recognition and Artificial Intelligence 19(5), 701–713 (2005)
8. Ma, J., Wang, L.: BYY Harmony Learning on Finite Mixture: Adaptive Gradient Implementation and A Floating RPCL Mechanism. Neural Processing Letters 24(1), 19–40 (2006)
9. Ma, J., Liu, J.: The BYY Annealing Learning Algorithm for Gaussian Mixture with Automated Model Selection. Pattern Recognition 40, 2029–2037 (2007)
10. Ma, J., He, X.: A Fast Fixed-point BYY Harmony Learning Algorithm on Gaussian Mixture with Automated Model Selection. Pattern Recognition Letters 29(6), 701–711 (2008)
11. Ma, J.: Automated Model Selection (AMS) on Finite Mixtures: A Theoretical Analysis. In: Proceedings of International Joint Conference on Neural Networks, Vancouver, Canada, pp. 8255–8261 (2006)
12. Ma, J., Xu, L., Jordan, M.I.: Asymptotic Convergence Rate of the EM Algorithm for Gaussian Mixtures. Neural Computation 12(12), 2881–2907 (2000)
13. Ma, J., Xu, L.: Asymptotic Convergence Properties of the EM Algorithm with respect to the Overlap in the Mixture. Neurocomputing 68, 105–129 (2005)
14. Ma, J., Fu, S.: On the Correct Convergence of the EM Algorithm for Gaussian Mixtures. Pattern Recognition 38(12), 2602–2611 (2005)
15. Zhang, Z., Chen, C., Sun, J., et al.: EM Algorithms for Gaussian Mixtures with Split-and-Merge Operation. Pattern Recogniton 36, 1973–1983 (2003)
16. Verbeek, J.J., Vlassis, N., Kröse, B.: Efficient Greedy Learning of Gaussian Mixture Models. Neural Computation 15(2), 469–485 (2003)